



Evidence for Excellence in Education

NFER Thinks

What the evidence tells us

Where have all the levels gone?

**The importance of a shared understanding of assessment
at a time of major policy change**



Plans to dismantle the National Curriculum levels are integral to the latest round of educational reforms – heralding one of the most fundamental changes to England’s assessment system for many years. At the same time, greater responsibilities for schools in tracking and assessing pupil progress form another part of the proposed agenda.

This paper, published in advance of NFER’s response to the Department for Education’s consultation on primary assessment and accountability, argues that while it may not be desirable to preserve the language of ‘levelness’, the importance of a shared understanding of assessment should never be overlooked. NFER calls on professional associations, advisers and schools to support novice and expert teachers with high-quality professional development in assessment.

A Department for Education (DfE) article states:

As part of our reforms to the national curriculum, the current system of 'levels' used to report children's attainment and progress will be removed. It will not be replaced.

(DfE, 2013a)

The context

Like it or not (and many did not), the National Curriculum levels have become embedded in the shared language of educational assessment in England. Throughout two decades of curriculum revisions, testing reforms and countless other education policy changes, they have persisted, working their way into the educational psyche.

They made their first appearance in the early stages of the statutory National Curriculum, back in 1988. In those days, the concept of levelling represented an 'innovative assessment system' (Whetton, 2009, p.141). In the TGAT (Task Group on Assessment and Testing) report (Department of Education and Science and the Welsh Office, 1988), a novel level-related structure was proposed as a national assessment framework. That report introduced the term destined to resonate through the decades in national standards debates: 'We shall use the word level to define one of a sequence of points on a scale to be used in describing the progress of attainment' (Department of Education and Science and the Welsh Office, 1988, p.32).

What was the point of levels? You could say that they fused the relationship between the curriculum and assessment, making possible the **criterion-referenced assessment** of pupil performance against that curriculum.

What is **criterion-referenced assessment**?

the measurement of students' performances against a set of criteria specifying educational attainments and ability levels.

(Kempa and L'Odiaga, 1984, p.56)

In the language of the TGAT report, levelling would allow pupil progress to be 'defined in terms of the national curriculum, and the stages of progress to be marked by levels of achievement as derived from that curriculum' (Department of Education and Science and the Welsh Office, 1988, p.30). Crucially, levels secured a move away from the then-dominant practice of **norm-referenced assessment**.

What is **norm-referenced assessment**?

The simplest statistical approach to maintaining standards is to ensure that the proportion of candidates awarded each grade remains the same from one examination to the next. This is most commonly referred to as norm referencing.¹

(Baird et al., 2000, p.220)

By 1995, there were eight levels specifying criteria that pupils needed to demonstrate – each so-called 'level descriptor' a step up in difficulty from the previous one. In its entirety, this measure set out, level by level, the incremental progression for a given curriculum subject.

The mapping of levels to the main stages of schooling gave us level 2 as the nationally expected standard of achievement at age seven (end of key stage 1) and level 4 as the benchmark for the majority at age 11 (end of key stage 2). The levels continued into secondary education: expected progression was, for the first time, set out and charted for pupils from the age of five. Level-based assessment outcomes were used for both **formative** and **summative** purposes including, of course, the end-of-key-stage tests. After some years of use, a commonly held view, at least in some parts of the education system, was 'You may or may not like them, but you know where you are with levels'.

What is the **formative use of assessment information**?

Formative is the use of day-to-day, often informal, assessments to explore pupils' understanding so that the teacher can best decide how to help them to develop that understanding.

(Mansell et al., 2009, p.9)

What is the **summative use of assessment information**?

Summative is the more formal summing-up of a pupil's progress that can then be used for purposes ranging from providing information to parents to certification as part of a formal examination course.

(Mansell et al., 2009, p.9)

But *did* you know where you were with levels? Not everyone would agree with that. From a conceptual perspective, NFER assessment specialists questioned the relationship between judgements made using National Curriculum levels and criterion-referencing (Sizmur and Sainsbury, 1997; Sainsbury and Sizmur, 1998). Then there are the well-documented discussions about how possible it is to make reliable and consistent judgements confirming that a pupil is working at a certain level (William, 2003; Newton, 2003, 2009).

There was also plenty of debate around so-called level 'equivalence' in relation to the award of the same level in different pupil age groups:

What, for example, does it mean to be working at level 3? At key stage 1, this would indicate high achievement. Key Stage 2 pupils who achieve level 3 in English will be categorised as performing below expectations for that year group, given 85% are expected to be level 4 or above. (Stobart, 2009, p.171)

¹ Baird et al. note, in relation to norm referencing, that 'a more appropriate name is cohort referencing (William, 1996a). In cohort referencing, each grade boundary (the minimum mark at which a grade is awarded) is fixed such that a predetermined proportion of candidates achieves each grade'. (Baird et al., 2000, p.220).

The current situation

It is clear that there were many reasons why levels were not unequivocally regarded as synonymous with best assessment practice.² But why, exactly, has the DfE decided to abandon them? The answer, it seems, is complexity:

We believe this system is complicated and difficult to understand, especially for parents. It also encourages teachers to focus on a pupil's current level, rather than consider more broadly what the pupil can actually do.
(DfE, 2013a)

The DfE's current consultation on *Primary assessment and accountability under the new curriculum* (DfE, 2013c) suggests a return to **norm-referenced assessment** for **summative assessment** purposes. It proposes 'decile ranking' for reporting the results of National Curriculum tests: showing 'each pupil's ranking in the national cohort by decile (i.e. 10% of the cohort)' (DfE, 2013c). Concerningly, it is difficult to see exactly how this new approach will be easier to understand. In essence, it dislocates the link between the reporting of achievement and the learning that has taken place.

Certainly, the relationship between the new National Curriculum in England framework document and the assessment of that curriculum will need teasing out: in the absence of levels, under the heading 'Attainment Targets' is the sentence: 'By the end of each key stage, pupils are expected to know, apply and understand the matters, skills and processes specified in the relevant programme of study.' (DfE, 2013b, p.16). The tracking and mapping of pupil progress towards this end-point is not further defined.

Opportunities and challenges

There are some potential positives to take from the situation too. Perhaps the most interesting part of *Assessing without Levels* is the suggestion of a greater role for schools in the **formative assessment** process:

Schools will be able to introduce their own approaches to formative assessment, to support pupil attainment and progression.
(DfE, 2013a)

The direction of travel intimates opportunity for schools to have greater involvement, control and input into the shape and structure of assessment. This is surely to be welcomed in a climate where teachers, all too often, may feel that they are passive recipients of assessment reform. Since the influential work of Black and William (1998) starting in the late 1990s, there has been no shortage of interest in the principles and practice of formative assessment. There are many examples of how teachers' use of good practice assessment contributes to supporting pupil progress. For instance, a 2011 Ofsted evaluation concluded that, under strong school leadership, an initiative involving teachers in assessing pupil progress 'was an important factor in pupils' rising achievement, particularly in English and mathematics' (Ofsted, 2011, p.5). A recent small-scale Australian study of formative assessment in primary science illustrates the connection between teaching, learning and assessment that teachers can forge in the classroom to create the right conditions for developing pupils' understanding (Loughland and Kilpatrick, 2013).

2 *The Framework for the National Curriculum. A report by the Expert Panel for the National Curriculum review* (DfE, 2011) includes critical review of the level-based approach.

Our proposition – for assessment to work, you need a shared understanding of standards

For any assessment reform to be successful there is a central point that must not be overlooked. It has to do with a shared understanding of assessment. This is more than shared language or terminology, although we suggest that common terms of reference are a helpful starting point. It has, at its heart, the orchestration of a set of good-practice assessment principles, a common interpretation of assessment terminology and a shared understanding of the assessment standards to be applied. This shared understanding should allow all those involved in the process a quality and precision of communication. It is only then that accurate judgements can be made and pupil progress supported. Ofsted found that good assessment materials

provided teachers with a common language to discuss and agree pupils' progress. This improved the consistency of assessment practice.
(Ofsted, 2011, p.5)

The level-based language of the National Curriculum was not perfect. But it did offer a certain degree of shared vocabulary and conceptualisation to aid communication about pupil progress. At its best, it informed assessment discussion and resulted in high-quality formative assessment practice. Teachers used a common interpretation of the criteria and exemplification of standards to benchmark their pupils' achievements, identify areas for development and plan the next steps. You could argue that good teachers don't need the language of 'levelness' to do this, and maybe that's so. But the principles of good-practice assessment tell us that it's important to have some shared point of reference for assessment standards. The alternative would risk a return to assessment localism in its worst sense: assessment with no agreed external reference point, leading to uncertainty about standards in pupil achievement.

Conclusions

NFER believes that a shared understanding of assessment is inextricably linked with teachers' professional development.

In order to introduce and maintain rigorous and consistent approaches to assessment in schools, novice and established teachers need to develop a culture and discourse of high-quality assessment throughout their careers. The role of professional associations,³ advisers and schools should be to support this development, so that a shared understanding of assessment can become embedded in classroom practice. This is something that NFER, as an independent educational charity, takes very seriously.⁴

Whatever the Government's plans for the future shape of assessment, one thing is certain: we can best support pupil learning and progression through a commitment to supporting the professional understanding of educational assessment.

3 The Association of Teachers and Lecturers (ATL) identifies the 'development, understanding and application of assessment criteria' as one of 'the key areas of professional development support required for teachers' (ATL, 2008).

4 NFER has published a series of free-to-download leaflets called *Getting to grips with assessment: Primary* for teachers and others interested in good assessment practice (NFER, 2007).

References

- Association of Teachers and Lecturers (2008). *Assessing to Learn: Teachers at the Heart of Assessment*. London: ATL [online]. Available: www.atl.org.uk/Images/Assessing%20to%20learn%20PS%202008.pdf [18 September 2013].
- Baird, J-A., Cresswell, M. and Newton, P. (2000). 'Would the real gold standard please step forward?', *Research Papers in Education*, **15**, 2, 213–229.
- Black, P. and Wiliam, D. (1998). *Inside the Black Box: Raising Standards Through Classroom Assessment*. London: GL Assessment.
- Department for Education (2011). *The Framework for the National Curriculum: a Report by the Expert Panel for the National Curriculum Review*. London: DfE [online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/175439/NCR-Expert_Panel_Report.pdf [18 September 2013].
- Department for Education (2013a). *Assessing Without Levels* [online]. Available: www.education.gov.uk/schools/teachingandlearning/curriculum/nationalcurriculum2014/a00225864/assessing-without-levels [18 September 2013].
- Department for Education (2013b). *The National Curriculum in England: Framework Document*. London: DfE [online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/210969/NC_framework_document_-_FINAL.pdf [18 September 2013].
- Department for Education (2013c). *Primary Assessment and Accountability Under the New National Curriculum*. London: DfE [online]. Available: https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/223672/Primary_assessment_and_accountability_under_the_new_curriculum_consultation_document_pdf.pdf [18 September 2013].
- Department of Education and Science and the Welsh Office (1988). *National Curriculum Task Group on Assessment and Testing: a Report*. London: Department of Education and Science [online]. Available: www.educationengland.org.uk/documents/pdfs/1988-TGAT-report.pdf [18 September 2013].
- Kempa, R. and L'Odiaga, J. (1984). 'Criterion-referenced interpretation of examination grades', *Educational Research*, **26**, 1, 56–64.
- Loughland, T. and Kilpatrick, L. (2013). 'Formative assessment in primary science', *Education 3-13: International Journal of Primary, Elementary and Early Years Education*, 1–13 (DOI: 10.1080/03004279.2013.767850) [online]. Available: www.tandfonline.com/doi/abs/10.1080/03004279.2013.767850 [18 September 2013].
- Mansell, W., James, M. and Assessment Reform Group (2009). *Assessment in Schools: Fit for Purpose? A Commentary by the Teaching and Learning Research Programme*. London: Teaching and Learning Research Programme [online]. Available: www.tlrp.org/pub/documents/assessment.pdf [18 September 2013].
- National Foundation for Educational Research (2007). *Getting to Grips with Assessment: Primary*. Slough: NFER [online]. Available: www.nfer.ac.uk/publications/getting-to-grips-with-assessment-primary/ [18 September 2013].
- Newton, P. (2003). 'The defensibility of national curriculum assessment in England', *Research Papers in Education*, **18**, 2, 101–127.
- Newton, P. (2009). 'The reliability of results from national curriculum testing in England', *Educational Research*, **51**, 2, 181–212.
- Ofsted (2011). *The Impact of the 'Assessing Pupils' Progress' Initiative*. Manchester: Ofsted [online]. Available: www.ofsted.gov.uk/sites/default/files/documents/surveys-and-good-practice/t/The%20impact%20of%20the%20Assessing%20pupils%20progress%20initiative.pdf [18 September 2013].
- Sainsbury, M. and Sizmur, S. (1998). 'Level descriptions in the national curriculum: what kind of criterion referencing is this?', *Oxford Review of Education*, **24**, 2, 181–193.
- Sizmur, S. and Sainsbury, M. (1997). 'Criterion referencing and the meaning of national curriculum assessment', *British Journal of Educational Studies*, **45**, 2, 123–140.
- Stobart, G. (2009). 'Determining validity in national curriculum assessments', *Educational Research*, **51**, 2, 161–179.
- Whetton, C. (2009). 'A brief history of a testing time: national curriculum assessment in England 1989–2008', *Educational Research*, **51**, 2, 137–159.
- William, D. (2003). 'National curriculum assessment: how to make it better', *Research Papers in Education*, **18**, 2, 129–136.

NFER is a charity with a reputation worldwide for providing independent and robust evidence for excellence in education and children's services. Our aim is to improve education and learning, and hence the lives of learners, by researching what is happening now. Our authoritative insights inform policy and practice, offering a unique perspective on today's educational challenges.

We work with an extensive network of organisations, all genuinely interested in making a difference to education and learners. Any surplus generated is reinvested in research projects to continue our work to improve the life chances of all learners from early years through to higher education.



Frances Brill

Dr Frances Brill is a senior research manager at NFER. Her main research expertise is literacy assessment in the primary age range. She has managed and contributed to a wide range of literacy and assessment research projects. These include the development of classroom materials

for speaking and listening teacher-assessments, National Curriculum tests in English and the provision of literacy consultancy to the DfE.

Frances has a specialist background in linguistics and education, holding an MA in general linguistics and a PhD in educational linguistics. Her teaching experience includes primary teaching at key stage 2, and lecturing in syntax to undergraduates.

Frances is the co-editor of NFER's peer-review international journal, *Educational Research*, which is listed in the Social Science Citation Index and published by Routledge. She is a fellow of The Chartered Institute of Educational Assessors and a practitioner of the Association for Educational Assessment – Europe.



Liz Twist

As head of the Centre for Assessment, Liz Twist oversees all assessment development and research projects at NFER, including national tests in English and science, and assessment research projects. She has directed a large number of assessment development projects, primarily in the area of literacy.

These include formal test development projects as well as investigations into assessment for learning.

Liz is national research coordinator for the Progress in International Reading Literacy Study (PIRLS), an international comparative survey of the reading skills of 9 to 10-year-olds.

Liz has a wide range of teaching experience including as deputy headteacher of a combined school; establishing and managing an integrated department for children with learning difficulties; and working in a secondary special school.

Published September 2013

© 2013 National Foundation for Educational Research
Registered charity no 313392

ISBN 978-1-908666-74-1

How to cite this publication:

Brill, F. and Twist, L. (2013). *Where Have All the Levels Gone? The Importance of a Shared Understanding of Assessment at a Time of Major Policy Change* (NFER Thinks: What the Evidence Tells Us). Slough: NFER.

National Foundation for Educational Research
The Mere, Upton Park
Slough, Berks SL1 2DQ

T: 01753 574123
F: 01753 691632
E: enquiries@nfer.ac.uk
www.nfer.ac.uk

